The report is very well written with properly structured sentences and paragraphs. The section 3.3 is especially very well written and I did not find too many things to edit there. Most of my edits were made for things before that. Also most of the changes suggested are related to changing the words in the topic position. I have edited some paragraphs too but it is totally alright if you find them unnecessary. They are mostly small changes to remove redundancy.

I also wanted to say that some of the sentences might be a little too long. I did not change all of them. I ended up editing just a few of the important ones.

1. For the second sentence, "Since we are depicting galaxy morphology by the aforementioned feature statistics and measuring cosmic time via redshift, galaxy morphology evolution is reflected in the change in conditional distribution of feature statistics given redshift.", you could instead change it to, "As mentioned earlier, the galaxy morphology is depicted by the aforementioned feature statistics and cosmic time is measured via redshift. Therefore galaxy morphology evolution is reflected in the change in conditional distribution of feature statistics given redshift."

2. In the next sentence instead of "there are several sets of feature statistics for each galaxy, i.e. correlated images taken in different filter bands", you could try "there are several sets of feature statistics for each galaxy, due to correlated images taken in different filter bands".

3. In the second sentence in 3.1, instead of " we would like to infer its full probability density function", you could replace it by "we infer its full probability density function" as it then makes infer the verb.

4. Next sentence, instead of "Thanks to some properties of probability distributions, we can achieve the goal with reasonable accuracy." you can try "We achieve the goal with reasonable accuracy due to some properties of probability distributions." That puts achieving the goal in topic position.

5. The next lines, "We can first recover the cumulative density function and then the probability density function. The specifics are as follows:" can either be totally killed as the algorithm itself says what you are doing or you can say "The following steps help us in recovering first the cumulative density function and then the probability density function."

6. In step 1, kill "so that get FZ(z)" after the equation.

7. In step 2, "get the probability density function, fZ(z) via differentiation:". Add the fZ(z).

8. Third sentence in 3.2, "If we assume the emitted light wavelength λemitted to be fixed (also known as the rest frame wavelength), there is a linear relationship between the redshift z and the observed light wavelength λobserved.", can be changed to "The redshift z and the observed light wavelength λobserved are linearly related if we assume fixed emitted light wavelength λemitted (also

known as the rest frame wavelength).” This gives importance to both redshift and observed wavelength and their linear relationship. You could put either λemitted or fixed in the stress position. I chose to put λemitted.

9. Change the bracket sizes in the third equation in 3.2.1

10. In “Assuming $\lambda = 4500A°$ as the rest-frame wavelength, the Y Band is associated with redshift range $z \in [1.05,1.63]$, $z \in [1.46,2.09]$ corresponds to the J band, and $z \in [2.12,2.71]$ the H Band.” its better to stick to symmetry. So I would suggest, “Assuming $\lambda = 4500A°$ as the rest-frame wavelength, the Y Band is associated with redshift range $z \in [1.05,1.63]$, the J band is associated with $z \in [1.46,2.09]$, and the H Band is associated with $z \in [2.12,2.71]$.”

11. Instead of the paragraph, “Using the photometric redshift measurement, we can compute the probability that $4500A°$ photons for a given galaxy are observed in each band, and select the band associated with highest probability if the probability if greater than 0.8 (or 0.6). Otherwise the galaxy is thrown out completely.” I would suggest, “For every galaxy, we compute the probability that $4500A°$ photons are observed in each band. If the band with the highest probability, has a probability of more than 0.8 (or 0.6), then we select the band or else we throw out the galaxy completely.”

12. In the next paragraph, first line “For galaxies in field GOODSN, we only have a point estimate for each galaxy's redshift. We would identify a galaxy with a redshift band if its redshift point estimate is within that band's redshift range. ” think of replacing by “For galaxies in field GOODSN, as we only have a point estimate for each galaxy's redshift. Hence we pick a redshift band for a galaxy, if the point estimate is within that band's redshift range. ” Adding the hence gives a continuity to how the previous sentence links to the next one.

13. In the same paragraph, “ 1/3” can be replaced by “a third of the”

14. Last line of the subsection, instead of “following table” give the table number. Also think of including the sentence in the previous paragraph itself.

15. The first sentence in 3.2.2 can be killed as it is implied in the paragraph following it. Also the “means” in “ redshift bin means that its 4500A” can be changed to “implies.”

16. Further the same paragraph can be made into, “As we mentioned in 3.2.1, a galaxy belongs to a redshift bin implies that its $4500A°$ light is mostly likely to be observed in the corresponding filter band. So for any galaxy, the morphological statistics taken in that filter would reflect the morphology at $4500A°$. Hence galaxy morphologies become comparable via their feature statistics taken in the corresponding filters. The three groups of feature statistics reflect the morphology of galaxies at $4500A°$ in redshift range [1.05,1.63],[1.46,2.09], [2.12,2.71].”

This helps in reducing redundant facts mentioned in later sentences that are implied by the previous ones.

17. Page 11, third line, " Noise is present of noise in the density function, leading to lots of spurious peaks in density function." Change this into, "The presence of noise leads to lots of spurious peaks in the density function."

18. "So we first smooth the density function by convolving it with a normal kernel and then identify the modes using derivative conditions." This repeats a lot of information which comes from the next steps too. So instead "So we first smooth the density function before identifying the modes by following the steps given below."

19. In 3.3.2, you can try making some of the sentences a little shorter by dividing them into multiple ones.

20. Page 12 last line, last word, you could use as instead of i.e. Just after the equation, instead of "where $\hat{f}(-i)h(Z(i))$ is the vector of predicted values for redshift values in the ith. portion using the model trained with data beyond this portion with bandwidth h, and $\sim Y(i)j$ is the vector of true values of the jth statistic in the ith portion." you can write before the equation that "The redshift values in the ith portion are predicted using a model which is trained of data beyond the portion using bandwidth h." Then the equation, followed by "where $f(-i)h(Z(i))$ is the vector of predicted values for redshift values in the ith portion, h is the bandwidth used and $Y(i)j$ is the vector of true values of the jth statistic in the ith portion."

21. The beginning of 3.2.3, can be made "We test for significance in the redshift bin comparisons using a bootstrap analog of the ANOVA F test described in Zhou and Wong (2011). We can not use any of the commonly used methods as the distribution of feature statistics include abundance of outliers and so the assumption of normality is violated. The null hypothesis of the test is that the feature statistics in different redshift bins come from the same underlying distribution and the alternative hypothesis is that at least one of the redshift bins differ significantly from the rest." and the first three paragraphs can be reduced to just one paragraph. Also the first line talks about the theme of the entire subsection while the paragraph talks about what are you testing and how.